

Besides the uncannily identical “Jefferson” in the presidents' names, the ceremony leaves no doubt that it was set up to be a photo opportunity and the president was bathing in the glory of the accomplishment. It is ironic that the same technique used to sequence the human genome was also used in a 1999 DNA test to prove that President Thomas Jefferson could have fathered a child with a slave Sally Hemmings. And here we recall the outcome of the 1999 DNA fingerprinting of a stain on the famous Monica Lewinsky's blue dress, an evidence in the very politically motivated allegation of Bill Clinton's White House infidelity.

We believe the White House ceremony will be a footnote to this momentous event. Though President Clinton will be remembered for many other great achievements during his tenure, his name will likely be forgotten in this context. The same can be said of his counterpart, Prime Minister Tony Blair of United Kingdom, who held in tandem a ceremony in London to underscore the international nature of the genome project. But the wind of time will never erase the line etched into the sand of history by the completion of the first draft of the human genome.

### 3 The Human Genome Tour

The human body has 10 trillion cells. With the exception of red blood cells, each cell contains the entire human genome, all the genetic information needed to “build” a human being.

The entire human genome has 22 pairs of chromosomes (autosomes), and the X and the Y chromosome (sex chromosomes). A male has 22 pairs plus an X and a Y. A female has 22 pairs plus two X's. The 23 pairs of chromosomes that make up the human genome contain the complete information for human development. The chromosomes themselves are largely made up of very long DNA molecule, which has been elaborately wound up. The DNA, if stretched out, will be about 5 feet long and 50 trillionths of an inch wide. Somewhere hidden in the DNA coils and strung out in definite order are genes. The genes are simply sections of a chemical message which runs along the spine of the DNA molecule. Four chemical bases or nucleotides: Adenine (A), Guanine (G), Cytosine (C) and Thymine (T), run the full length of the DNA molecule. The order of these chemicals constitutes a code, instructing cells to make different proteins. The total information stored in the genome, therefore, can be most simply thought of as a very large encyclopedia containing the recipe for life. Despite the complexity of a human being, the recipe is written with an alphabet of only the four letters, A, C, G, and T. Using this analogy, *Encyclopedia Humanica* is three billion letters long from end to end. Somewhere in the encyclopedia are the clues to genetic diseases and phenotypes.

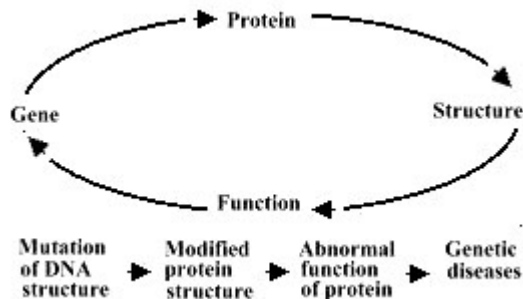
Pushing the encyclopedia analogy further, *Encyclopedia Humanica* contains all of the instructions needed to go from a fertilized egg to a complete human being. In this analogy, *Encyclopedia Humanica* has 22 volumes, and volume X and volume Y.

Much of what is found in *Encyclopedia Humanica* is very much similar to the encyclopedia of the mouse, for example. Many sentences that appear in a human's DNA also appear in a mouse's DNA. This reflects the fact that the two texts were the same until some point in the past when variations were introduced and the two species diverged from their common ancestry. In fact, the best way to read *Encyclopedia Humanica* is to read it alongside other encyclopedias, such as those of the mouse, fruit fly or nematode. Fruit flies are good model organisms for they reproduce in large quantities, and their life cycle is short. Besides ethical issues associated with experimenting with humans, short life cycle is a good attribute for experimentation with effects that can be passed on (inherited), and reproducing in large number is an advantage for statistical analysis. For experimentation purposes, humans reproduce too slowly and too few in numbers.

Genetic variations in genomes, various possible combinations of different possible genes, like different spellings of the same word, create the infinite variety that we see among individuals of a species.<sup>3</sup> For example, between the genomes of

- Bart and Lisa Simpson, two blood siblings, there is a difference in 2 million base pairs (bp)
- Tarzan and Jane, two individuals of the same species, there is a difference of 6 million bp in their genomes.
- Tarzan and chimpanzee, two organisms of closely related species, there is a difference of 50 million bp in their genomes.
- Popeye and spinach, two organisms of distant species, there is a difference of 2 billion bp in their genomes.

Besides these genotype differences, there are also other variations. Genetic variations lead to diseases:



**Figure 2.** Genes code for proteins. Proteins have particular structures leading to different functions. A mutated DNA can lead to a modified protein structure, thus abnormal protein function. This is manifested as abnormalities or diseases.

<sup>3</sup> Hwa A. Lim, "Lecture Note at Science & Computer Teachers Summer Training", Supercomputer Computations Research Institute, Tallahassee, Summer 1992.